Vlerick Leuven Gent Management School

the Autonomous Management School of
Ghent University and Katholieke Universiteit Leuven

Vlerick Leuven Gent Working Paper Series 2006/29

A NOTE ON PERFORMANCE MEASURES FOR FAILURE PREDICTION MODELS

HUBERT OOGHE
Hubert.Ooghe@vlerick.be
CHRISTOPHE SPAENJERS

D/2006/6482/31

A NOTE ON PERFORMANCE MEASURES FOR FAILURE PREDICTION MODELS

HUBERT OOGHE

Vlerick Leuven Gent Management School

CHRISTOPHE SPAENJERS

Ghent University

Contact:

Hubert Ooghe

Vlerick Leuven Gent Management School

Tel: +32 09 210 97 86

Fax: +32 09 210 97 00

Email: Hubert.Ooghe@vlerick.be

1. INTRODUCTION

Since decades, the topic of business failure prediction has been an important research area for both academics and practitioners. Bankruptcy prediction involves the classification of firms in a failing and a non-failing group[1]. Generally, this classification is based on (1) a prediction model that attributes a 'score' to each firm in the data set and (2) a certain cut-off point. To evaluate the classification results, several performance measures can be used. This note outlines these measures and illustrates the connections between them with numerical examples. This may help the reader to better understand (and possibly use) these classification measures[2].

2. CLASSIFICATION AND ERROR RATES

In most research papers on failure prediction, statistical techniques like multiple discriminant analysis or logit analysis are applied[3]. Generally, a model score is calculated for each firm / observation in the data set. When classifying observations in two mutually exclusive groups, a cut-off point has to be chosen. We assume here that a company will be classified as 'non-failing' if its score is higher than the cut-off point and classified as 'failing' if it is lower[4].

Two types of misclassifications can then be made. A type I error represents a 'credit risk': a failing firm is classified as a non-failing one. A type II error represents a 'commercial risk': a non-failing firm is classified as a failing one. Both errors come with its costs. Altman (1980) mentions different components of type I and type II costs in the context of commercial bank lending.

[1] Other classifications are also possible, e.g. 'bankrupt' versus 'non-bankrupt' or 'financially distressed' versus 'not distressed'.
[2] Of course, the use of these measures is not limited to bankruptcy prediction.
[3] See Balcaen & Ooghe (2006) for an overview of much-used statistical methodologies.
[4] This can also be the other way around, depending on the construction of the model.

3. UER, D-MAX AND GINI-COEFFICIENT

The determination of an optimal cut-off point is not as easy as might be expected[5]. We hereafter assume that the loss functions of type I and type II errors are symmetrical. Although the type II error only leads to an opportunity cost, it is not incredible that this cost is as high as the more visible cost that goes with a type I error. Furthermore, as noted in Ooghe et al. (2005), "*the allocation of weights to the different types of errors is subjective and depends on the degree of risk aversion of the risk analyst*". We also do not take into account the population proportions[6]. To determine the optimal cut-off point, we have to minimize the unweighted average of the type I and type II error rates, or **UER** (unweighted error rate).

At each possible cut-off point *c*, the type II error can be measured by the cumulative distribution function of the non-failing firms (F_{nf}). This function gives the percentage of non-failing firms that have a score smaller than (or equal to) the cut-off point, and thus are misclassified as failing firms. Analogously, the type I error is equal to 1 minus the value of the cumulative distribution of the failing firms ($1 - F_f$), since this gives the percentage of failing firms with a score higher than the cut-off point *c*.

The optimal cut-off point – at which the unweighted average of the two types of error rates is minimal – is also the point at which there is the largest difference *D* (the so-called **D-max**) between the cumulative distribution functions F_f and F_{nf}. The D-max is the central statistic of the Kolmogorov-Smirnov two-sample test (Siegel and Castellan, 1988).

D-max is equal to $max\ [F_f - F_{nf}]$, while the minimum UER can be expressed as $min\ [(1 - F_f + F_{nf}) / 2]$. Also, $D = 1 – 2 * UER$ at each cut-off point. The cut-off point with the lowest UER thus corresponds to the score that discriminates most between failing and non-failing firms.

A model can also be evaluated on its power to discriminate between failing and non-failing firms not only at the optimal cut-off point, but at each possible cut-off point. We then evaluate the performance of a model based on the "inequality principle" (Joos et al., 1998), which means that we measure the aggregate inequality of the two distributions (failing and non-failing). We do so by constructing a trade-off function of the two types of error rates.

The graph of this trade-off function thus plots all possible combinations of type I and type II error rates, i.e. the type I and type II errors at each possible cut-off point. The type II error rate (F_{nf}) is situated on the X-axis, while the Y-axis gives the corresponding type I error

---

[5] Hsieh (1993) and Koh (1994) discuss some of the difficulties of determining an optimal cut-off point.

rate (1 - F_f). The closer the trade-off function is situated to the axes, the more the model discriminates between failing and non-failing firms. The 'best' possible model is the one that coincides with the two axes. There one can choose every possible combination of type I and type II error rates, including both times 0%. The 'worst' possible model does not discriminate and is a trade-off function between the two types of error rates, but the sum of the two is always equal to 100%.

The **Gini-coefficient**, a measure for the discriminating power of a model, can then be calculated as the area between the trade-off function of the model in question and the trade-off function of this worst, non-discriminating model, divided by the area between the trade-off functions of the best and the worst model. We thus get a coefficient between 0 and 1. It is important to see that we do not have to calculate an optimal cut-off point here.

Based on Joos et al. (1998), we can give the following empirical approximation of the Gini-coefficient: $\hat{GINI} = 1 - \sum_{i=1}^{n}(x_i - x_{i-1})(y_{i-1} + y_i)$ with x_i and y_i equal to the type II and type I error when using cut-off point i.

4. SOME NUMERICAL EXAMPLES

In this section we give some hypothetical examples, by which we can illustrate the use of the performance measures mentioned above. Models 1 to 4 are assumed to be failure prediction models that attribute a score between 0 and 1 to every firm (observation).

Insert Model 1 About Here

Models 1 and 2 are extremes. In the first model, one can discriminate perfectly between failing and non-failing firms by making the cut-off point c equal to 0,50. All failing firms have a model score lower than 0,50, while all non-failing firms score higher. At the cut-off point, the unweighted error rate equals 0%, and the D-max between the two cumulative distributions is 100%. As a consequence, the Gini-coefficient is 1. Also in the graph of the trade-off function (at the end of this section) it becomes clear that one can choose for a cut-off point at which both error rates are equal to 0%. Therefore, model 1 is the best possible model.

[6] See Joos et al. (1998) for more information on the impact of population proportions and misclassification costs.

Insert Model 2 About Here

The second model is the worst possible model. It does not discriminate between failing and non-failing firms, since the cumulative distribution functions coincide. One can not select a cut-off point at which more failing than non-failing firms score lower. The UER is thus 50% at each possible *c*. The Gini-coefficient is zero. The trade-off function goes from 100% type I error rate (and a type II error rate of 0%) to 100% type II error rate (and a type I error of 0%).

Insert Model 3 About Here

In model 3, we can determine two optimal cut-off points. Both have a difference between the cumulative distributions of 40% and an average error rate of 30%. The difference is in the proportions of type I and type II errors. Following our assumptions, both cut-off points are equally valuable. The Gini-coefficient can be calculated as the area between the model (model 3) and the worst model (model 2), scaled by the area between the best model (model 1) and the worst model (model 2). In this case the Gini-coefficient is 0,360.

Insert Model 4 About Here

The fourth model gives a more realistic view of a failure prediction model. 78% of the firms have a score lower than or equal to 0,40, while 87% of the non-failing firms score higher. We thus get a UER of 17,5% and a D-max of 65%. The discriminating power of the model – as measured by the Gini-coefficient (0,773) – is of course a lot higher than that of model 3.

Insert Trade-off functions models 1-4 About Here

In this graph we see the ‘ideal’ model 1 coinciding with the axes, the non-discriminating model 2 going from 100% type I error to 100% type II error, and models 3 and 4 in between. The circles indicate the combination of error rates at the optimal cut-off points.

5. CONCLUSION

In this note we briefly described some important performance measures that can be used in failure prediction research. We start from a failure prediction model that attributes a score from 0 to 1 to each firm, where a higher score indicates a lower chance of failure. Assuming equal misclassification costs and equal population proportions, an optimal cut-off point can be calculated by minimizing the unweighted average of the type I and type II error rates (UER). At this cut-off point, the difference between the cumulative distributions of the failing and the non-failing firms will reach its maximum (D-max). With a Gini-coefficient one can measure the total discriminating power of a model.

REFERENCES

Altman, E.I., 1980, Commercial bank lending: process, credit scoring, and costs of errors in lending, Journal of Financial and Quantitative Analysis, Vol. 15, nr. 4, November 1980, p. 813-831

Balcaen, S. and Ooghe, H., 2006, 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems, British Accounting Review, Vol. 38, nr. 1, p. 63-93

Hsieh, S., 1993, A note on the optimal cutoff point in bankruptcy prediction models, Journal of Business Finance & Accounting, Vol. 20, nr. 3, April 1993, p. 457-464

Joos, P., Ooghe, H. and Sierens, N., 1998, Methodologie bij het opstellen en beoordelen van kredietclassificatiemodellen, Tijdschrift voor Economie en Management, Vol. 18, nr. 1, p. 3-48

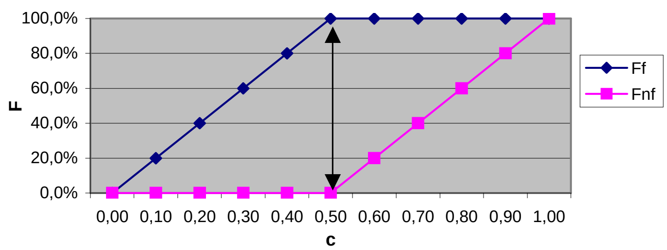
Koh, H.C., 1992, The sensitivity of optimal cutoff points to misclassification costs of Type I and Type II errors in the going-concern prediction context. Journal of Business Finance & Accounting, Vol. 19, nr. 2, January 1992, p. 187-197

Ooghe, H., Spaenjers, C. and Vandermoere, P., 2005, Business failure prediction: simple-intuitive models versus statistical models, Paper nr. 05/338, Working paper series, Faculty of Economics and Business Administration, Ghent University, Belgium, 37 pp.

Siegel, S. and Castellan, N.J., 1988, Nonparametric statistics for the behavioral sciences, McGraw-Hill, New York, 2nd edition

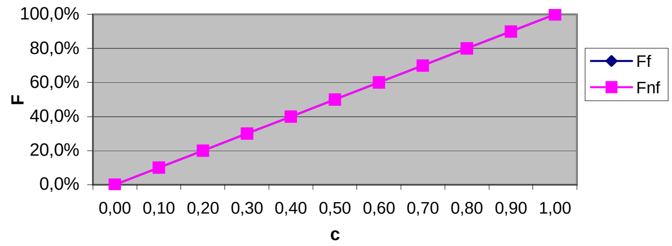
MODEL 1 PERFECTLY DISCRIMINATING

c	Ff	1 - Ff	Fnf	UER	D	Gini
0,00	0,0%	100,0%	0,0%	50,0%	0,0%	
0,10	20,0%	80,0%	0,0%	40,0%	20,0%	0,000
0,20	40,0%	60,0%	0,0%	30,0%	40,0%	0,000
0,30	60,0%	40,0%	0,0%	20,0%	60,0%	0,000
0,40	80,0%	20,0%	0,0%	10,0%	80,0%	0,000
0,50	**100,0%**	**0,0%**	**0,0%**	**0,0%**	**100,0%**	0,000
0,60	100,0%	0,0%	20,0%	10,0%	80,0%	0,000
0,70	100,0%	0,0%	40,0%	20,0%	60,0%	0,000
0,80	100,0%	0,0%	60,0%	30,0%	40,0%	0,000
0,90	100,0%	0,0%	80,0%	40,0%	20,0%	0,000
1,00	100,0%	0,0%	100,0%	50,0%	0,0%	0,000
						1,000



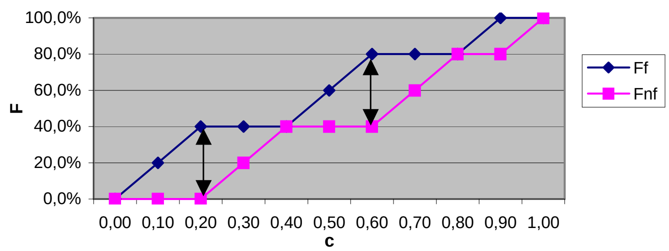
MODEL 2 (NON-DISCRIMINANT)

c	Ff	1 - Ff	Fnf	UER	D	Gini
0,00	0,0%	100,0%	0,0%	50,0%	0,0%	
0,10	10,0%	90,0%	10,0%	50,0%	0,0%	0,190
0,20	20,0%	80,0%	20,0%	50,0%	0,0%	0,170
0,30	30,0%	70,0%	30,0%	50,0%	0,0%	0,150
0,40	40,0%	60,0%	40,0%	50,0%	0,0%	0,130
0,50	50,0%	50,0%	50,0%	50,0%	0,0%	0,110
0,60	60,0%	40,0%	60,0%	50,0%	0,0%	0,090
0,70	70,0%	30,0%	70,0%	50,0%	0,0%	0,070
0,80	80,0%	20,0%	80,0%	50,0%	0,0%	0,050
0,90	90,0%	10,0%	90,0%	50,0%	0,0%	0,030
1,00	100,0%	0,0%	100,0%	50,0%	0,0%	0,010
						0,000



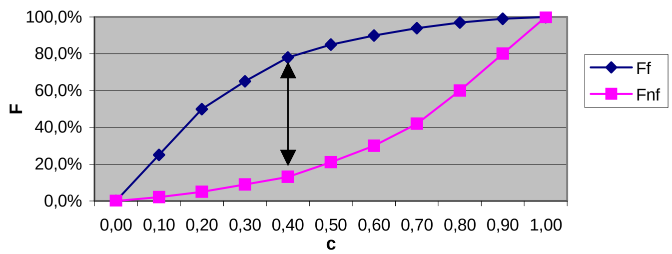
MODEL 3

c	Ff	1 - Ff	Fnf	UER	D	Gini
0,00	0,0%	100,0%	0,0%	50,0%	0,0%	
0,10	20,0%	80,0%	0,0%	40,0%	20,0%	0,000
0,20	**40,0%**	**60,0%**	**0,0%**	**30,0%**	**40,0%**	0,000
0,30	40,0%	60,0%	20,0%	40,0%	20,0%	0,240
0,40	40,0%	60,0%	40,0%	50,0%	0,0%	0,240
0,50	60,0%	40,0%	40,0%	40,0%	20,0%	0,000
0,60	**80,0%**	**20,0%**	**40,0%**	**30,0%**	**40,0%**	0,000
0,70	80,0%	20,0%	60,0%	40,0%	20,0%	0,080
0,80	80,0%	20,0%	80,0%	50,0%	0,0%	0,080
0,90	100,0%	0,0%	80,0%	40,0%	20,0%	0,000
1,00	100,0%	0,0%	100,0%	50,0%	0,0%	0,000
						0,360



MODEL 4

c	Ff	1 - Ff	Fnf	UER	D	Gini
0,00	0,0%	100,0%	0,0%	50,0%	0,0%	
0,10	25,0%	75,0%	2,0%	38,5%	23,0%	0,035
0,20	50,0%	50,0%	5,0%	27,5%	45,0%	0,038
0,30	65,0%	35,0%	9,0%	22,0%	56,0%	0,034
0,40	**78,0%**	**22,0%**	**13,0%**	**17,5%**	**65,0%**	0,023
0,50	85,0%	15,0%	21,0%	18,0%	64,0%	0,030
0,60	90,0%	10,0%	30,0%	20,0%	60,0%	0,023
0,70	94,0%	6,0%	42,0%	24,0%	52,0%	0,019
0,80	97,0%	3,0%	60,0%	31,5%	37,0%	0,016
0,90	99,0%	1,0%	80,0%	40,5%	19,0%	0,008
1,00	100,0%	0,0%	100,0%	50,0%	0,0%	0,002
						0,773



TRADE-OFF FUNCTIONS MODELS 1 - 4